MÔ TẢ

Công nghệ sử dụng:

Thư viện sử dụng:

Pyvi: Thư viện NLP tiếng việt

Tqdm:

Numpy

Gensim: Thư viện NLP

Os

Pathlib

Pickle: lưu file dạng byte

Sklearn

Tensorflow

Các mô hình đang sử dụng:

Logistic Regression:

Long short term memory (LSTM):

Bidirectional Recurrent Neural Network (BRNN):

Các bộ dữ liệu sau khi xử lí:

Đầu tiên sẽ là tách bộ luật theo các điều khoản

Với các điều khoản sẽ được xử lí text theo thư pyvi để biến text thành các từ có liên quan sẽ thành 1 cụm để xử lí

Sau đó sử dụng 1 số phương pháp tách text thành số:

Count Vectors

TF-IDF vectors:

Word Level

Ngram\_level

Sử dụng thuật toán SVD import từ thư viện để giám số lượng chiều nhưng mà vẫn đảm bảo được các yếu tố gốc

Tổ chức file/folder:

2 folder cha:

TEST\_FULL

TRAIN\_FULL

Trong mỗi folder cha sẽ có folder con tương đương với là nhãn của văn bản mà các folder con chứa ví dụ:

TEST\_FULL:

Hình Sự

Dân Sự

Tố tụng dân sự

Tố tụng hình sự

Hàng hải

Lao động

TRAIN\_FULL:

Hình Sự

Dân Sự

Tố tụng dân sự

Tố tụng hình sự

Hàng hải

Lao động

Training Model:

Logistic Regression:

Dự đoán đúng 97%

Thời gian chạy: nhanh

Trong trường hợp đata to hơn chưa chắc tốt

Long short term memory (LSTM):

Dự đoán đúng 88%

Thời gian chạy: nhanh

Cần điều chỉnh lại tập train để cải thiện

Bidirectional Recurrent Neural Network (BRNN):

Dự đoán đúng 85%

Thời gian chạy: lâu hơn so với LSTM

Cần điều chỉnh lại tập train để cải thiện

Xử lí file:

Tạo 6 Thư Mục Cho 6 Bộ Luật - Chúng ta có 6 bộ luật khác nhau. Mỗi thư mục sẽ chứa các tài liệu liên quan đến một bộ luật cụ thể. Đặt tên cho mỗi thư mục theo tên của từng bộ luật, ví dụ: “Bộ Luật 1”, “Bộ Luật 2”, v.v.

459734hỏ. Mỗi file văn bản có kích thước không quá 1024 ký tự. Điều này giúp dễ dàng quản lý và tìm kiếm thông tin trong tài liệu.

Sắp Xếp Các File Văn Bản - Mỗi file văn bản nên được sắp xếp theo định dạng sau: + Tên Bộ Luật: Ví dụ, “Bộ Luật Dân Sự”. + Ngày Tháng: Ngày và tháng mà bộ luật có hiệu lực, ví dụ “Ngày 01 tháng 01 năm 2024”. + Số Bộ Luật: Số hiệu hoặc tên của bộ luật, ví dụ “Số: 91/2015/QH13”. + Nội Dung Các Điều Khoản: Tên của các điều khoản hoặc điều, ví dụ “Điều 1: Quy định chung”.

1. Test Chatbot: (Đã ghi ở trong file Báo cáo sử dụng model) xây dựng chat bot dự đoán luật